# A time–frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses

**Article** *in* Pattern Recognition Letters · June 2011

DOI: 10.1016/j.patrec.2011.02.022 · Source: DBLP

CITATIONS
39

READS
449

**1 author:**

Hubert Cecotti
California State University, Fresno
**128** PUBLICATIONS **1,749** CITATIONS

SEE PROFILE

Pattern Recognition Letters 32 (2011) 1145–1153

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# A time–frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses

Hubert Cecotti *

*Institute of Automation (IAT), University of Bremen, Otto-Hahn-Allee, NW1, 28359 Bremen, Germany*

ARTICLE INFO

*Article history:*
Received 9 February 2010
Available online 11 March 2011
Communicated by R.C. Guido

*Keywords:*
Neural network
Convolution
Fourier transform
Spatial filters
Steady-state visual evoked potential (SSVEP)
Electroencephalogram (EEG)

ABSTRACT

A new convolutional neural network architecture is presented. It includes the fast Fourier transform between two hidden layers to switch the signal analysis from the time domain to the frequency domain inside the network. This technique allows the signal classification without any special pre-processing and uses knowledge from the problem in the network topology. The first step allows the creation of different spatial and time filters. The second step is dedicated to the signal transformation in the frequency domain. The last step is the classification. The system is tested offline on the classification of EEG signals that contain steady-state visual evoked potential (SSVEP) responses. The mean recognition rate of the classification of five different types of SSVEP response is 95.61% on a time segment length of 1 s. The proposed strategy outperforms other classical neural network architecures.



## 1. Introduction

Pattern recognition techniques are widely used for the classification and the detection of specific brain responses. Most of the effective solutions use machine learning techniques (Blankertz et al., 2006; Lotte et al., 2007; Müller et al., 2008). Almost all kinds of models have been used, like neural networks (Anderson et al., 1995; Cecotti and Gräser, 2008; Felzer and Freisieben, 2003; Haselsteiner and Pfurtscheller, 2000; Masic et al., 1995), support vector machines (SVM) (Blankertz et al., 2002; Rakotomamonjy and Guigue, 2008) and hidden Markov models (Obermaier et al., 2001; Zhong and Gosh, 2002). A steady-state visual evoked potential (SSVEP) response reflects the users attention to an oscillating visual stimulus. Flickering lights at different frequencies are usually used as stimuli. Their responses in the visual cortex correspond to SSVEP at the same frequencies and higher harmonics (Müller-Putz et al., 2005). These brain responses can be used for creating a Brain–Computer Interface (BCI) (Wolpaw et al., 2002; Sejnowski et al., 2007). BCIs based on the detection of SSVEP response have been used for neuroprosthetic device control, for the restoration of the grasp function in spinal cord injured persons (Müller-Putz and Pfurtscheller, 2008) and rehabilitation robot control (Lüth et al., 2007). The performance of a BCI is usually given by its information transfer rate and the detection accuracy of the commands.

Feed-forward neural networks with deep architectures has caught the attention of the pattern recognition community thanks to their performance in new empirical and theoretical studies, like for dimensionality reduction (Hinton and Salakhutdinov, 2006; Hinton et al., 2006) and the MNIST data set of handwritten Latin digits (Bengio et al., 2007). The knowledge of the recognition problem can provide information for restricting and tuning the neural network topology. Several convolutional neural networks (CNN) have been proposed for different problems like handwriting character recognition (LeCun et al., 1998a; Simard et al., 2003) and object recognition (LeCun et al., 2004). These CNNs provide high performance, allowing automatic feature extraction within their layers. It relatively conserves the input as raw data, except for scaling and centering the input vector. This strategy has many advantages: when the input data contain an inner structure, like for images and where invariant features shall be discovered. One goal is to avoid hand designed input features, which are not directly derivated by the problem, for decoding SSVEP response.

A new CNN architecture is proposed for classifying SSVEP responses based on electroencephalogram (EEG) signal. As the EEG signal contains a lot of variations, a classifier based on a CNN seems to be a good solution for the classification of SSVEP responses. The lack of robustness in static neural network classifiers with respect to time alignment of processed patterns suggests the need for a temporal or dynamic classifier in EEG classification (Barreto et al., 1996). The interest of the CNN is to directly classify the raw signal and to

* Present address: Department of Psychology and Institute for Collaborative Biotechnologies, University of California Santa Barbara, Santa Barbara, CA 93106-9660, USA. Tel.: +49 421 2183580; fax: +49 421 2184596.
*E-mail address:* hub20xx@hotmail.com

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.02.022

integrate the signal processing functions within the discriminant steps when it is needed. Indeed, it is not always possible to know the type of features to extract. However, high level features for the detection of SSVEP responses are known to be in the frequency domain, i.e. the frequency of the visual stimuli. Hence, it introduces new challenges for CNN for processing the data through different domains. The method that is proposed thereafter enables all the processing steps (preprocessing, feature extraction and classification) to be performed and tuned in a single and unified way.

The paper is organized as follows. Section 2 presents the problem of spatial filtering. The classifier architecture is described in Section 3. Section 4 is dedicated to the materials and the protocol experiment. Finally, the results and their discussion are detailed in the last two sections.

## 2. Spatial filtering

Spatial filtering, i.e. the creation of virtual electrodes, is one first step towards the classification of the EEG signal. Its goal is to enhance a particular information that is contained in the signal. Some electrodes may contain the same kind of noise that their combination may eliminate. Some electrodes contain more information than noise, a weight on their information power shall translate this behavior. Usually different spatial filters are created to perform the classification, they correspond to different channels. We distinguish the electrode inputs, which are a specific case of channels and channels that represent a combination of the electrodes, i.e. virtual electrodes. We define a channel as a linear combination of the signals measured by the $N_{elec}$ electrodes. A channel $c$ is defined by $N_{elec}$ weights. At each time $j$ the output value of a channel $c$ is:

$$c_j = \sum_{i=1}^{N_{elec}} w_i I_{i,j} \quad (1)$$

where $I$ is a 2-dimensional signal, $0 \leqslant i < N_{elec}$.

For basic channels, each channel corresponds to an electrode (if $i = j$ then $w_i = 1$ else $w_i = 0$). The creation of channels allows the analysis of a set of spatially independent vectors. The information from the electrodes is resumed in one scalar at a time $j$. For the EEG signal, a step is to find an optimal set $w(k)_i$, $0 \leqslant i < N_{elec}$, $1 \leqslant k < N_s$ where $N_s$ is the number of channels.

A channel usually represents the effect of a spatial filter. We distinguish three approaches for setting spatial filters:

- Fixed filters. For these filters, the weights are fixed manually. With the average combination, each electrode has the same weight. It assumes that the sinusoids have equal phases and the noise is equally distributed across the electrodes for a low dependence. However, the different electrode locations on the scalp can involve difference in the phases of the SSVEP responses (Burkitt et al., 2000). A better signal can be obtained by canceling the common nuisance signals with the bipolar approach (Müller-Putz et al., 2005). The Laplacian combination is an alternative to the bipolar solution. For instance, one electrode has a high weight whereas its neighboring electrodes have negative weights.
- Adaptive filters. These methods usually use statistical methods. Independent Component Analysis and Common Spatial Pattern (CSP) (Blankertz et al., 2008a,b; Brunner et al., 2007; Pfurtscheller et al., 1999; Tomioka et al., 2006) are often used. For instance, the CSP method is based on the second order statistics of the signal between electrodes. The filters are obtained by solving a generalized eigenvalue problem.
- Filters set with a generative approach. Such filters are set as a function of the expected signal to detect. For instance, the goal of the minimum energy combination (MEC) is to form combinations of the electrode signals, which cancels as much of the nuisance signals as possible. This technique removes any potential discriminant components from all the electrode signals, by projecting them onto the orthogonal complement of a formal model of the signal (Friman et al., 2007). It is based on the principal component analysis (PCA). With the maximum contrast combination, the relevant information in the signal is maximized and the energy in the nuisance signals is minimized simultaneously (Friman et al., 2007).

We propose a discriminant approach for classifying EEG signals, which correspond to different kinds of SSVEP responses. It allows creating spatial filters in a transparent way, i.e. spatial filters are not precisely defined as such. They are tailored together in relation to their discriminant power once they are combined during the classification. The goal is to determine the optimal set of weights for $N_s$ channels, which can improve the final classification.

## 3. Classifier overview

The classifier is based on a convolutional neural network (CNN), which is based on a multi-layer perceptron (MLP) with a special topology. Contrary to other CNN models described for handwritten

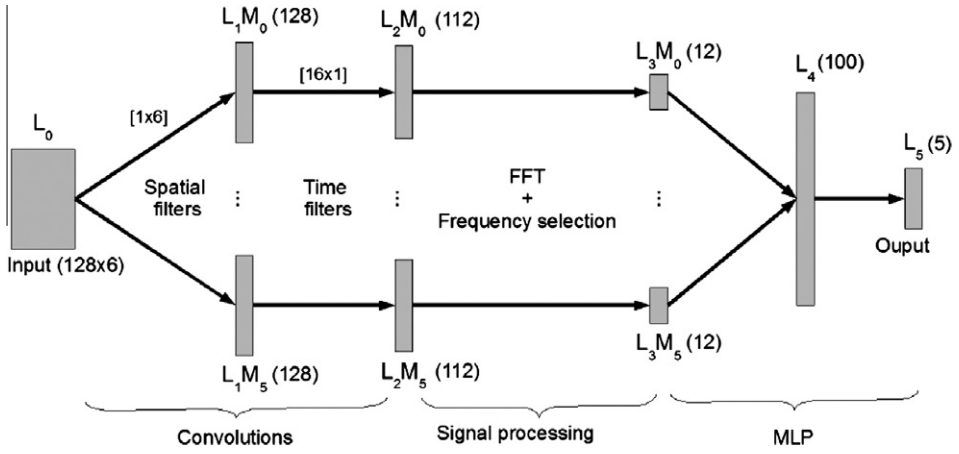


**Fig. 1.** System overview.

character recognition, the domain changes between the raw input features and the high level features for classifying SSVEP responses. This is the reason why we propose adding some signal processing methods between two hidden layers. For the SSVEP detection, the input is in the time-spatial domain, and the high level features are in the frequency domain. Indeed, we have to consider only the amplitudes of the SSVEP responses. The phase cannot be used as a feature as the phase of the visual stimuli are unknown. The high level features in the frequency domain must be therefore shift invariant. One problem is creating an architecture that solves the different signal processing problems during the classification. Fig. 1 displays an overview of the system. The different parts are all included in the neural network.

The information processing can be decomposed into three parts:

1. Data processing in the time domain: the neuron values have a semantic in the time domain. The network topology translates the application of the linear spatial and time filters. $N_{elec}$ and $N_k$ are the size of the spatial and time filters respectively.
2. The transfer from the time domain to the frequency domain: this step transfers the neuron values in the frequency domain. We consider $N_f$ frequencies, which are specified in relation to the frequencies of the stimuli. This frequency pool corresponds to the natural frequencies of the stimuli and their harmonics. As the phase of the signal is unkown, we have to provide to the classifier features in the frequency domain. In the time domain, the classifier should be shift invariant to correctly process and classify the signal.
3. Data processing in the frequency domain: the neuron values have a semantic in the frequency domain; they correspond to amplitude values of the signal. The number of classes in the problem is $M$, $M = N_{freq}$. Each class corresponds to a visual stimulus involving an SSVEP response.

### 3.1. Input

For SSVEP stimuli, we consider $N_{freq}$ visual stimulations on an LCD screen with boxes flickering at different frequencies. For the response, we consider $N_{elec}$ electrodes for the EEG signal acquisition. The signal corresponds to the voltage measure (in μV (microvolt)) between a reference electrode, and one of the $N_{elec}$ electrodes. The signal contains nuisance signals that have several origins: the environment, natural physical disturbances like other brain processes, breathing artifacts and involuntary muscular contractions.

We consider as input of the CNN a matrix $I$ of size $N_{elec} \times N_t$ where $N_t$ is the number of points that are considered for the analysis: $N_t = Fs * TS$. $N_t$ corresponds to the number of recorded samples in $TS$s with a sampling rate at $Fs$Hz. Before the classifier step, each signal sample is normalized independently for each sensor as to have a zero mean and standard deviation equal to one.

### 3.2. Neural network topology

The network is composed of six layers. Each layer is composed of one or several maps. We define a map as a layer entity that has a specific semantic: each map of the first hidden layer is a channel. The first hidden layer is dedicated to the creation of the different channels; the second hidden layer denoises the signal.

The network topology is described as follows:

- Layer 0 ($L_0$): the input layer. $I_{i,j}$ with $0 \leqslant i < N_{elec}$ and $0 \leqslant j < N_t$.
- Layer 1 ($L_1$): the first hidden layer is composed of $N_s$ maps. We define $L_1M_m$, the map number m. Each map of $L_1$ has the size $N_t$. This layer corresponds of the $N_s$ channels.
- Layer 2 ($L_2$): the second hidden layer is composed of $N_s$ maps. Each map of $L_2$ has the size $N_t - N_k$. Each map of $L_2$ is connected to its corresponding map in $L_1$.
- Layer 3 ($L_3$): the third hidden layer is composed of $N_s$ maps. Each map of $L_3$ has $N_f$ neurons.
- Layer 4 ($L_4$): the fourth hidden layer is composed of 1 map of 100 neurons. The size of this layer was chosen based on trial runs. This map is fully connected to the different maps of $L_3$.
- Layer 5 ($L_5$): the output layer. This layer has only one map of $M$ neurons, which represents $M$ frequencies to detect. This layer is fully connected to $L_4$.

### 3.3. Propagation

The value of a neuron in the layer $l$, in the map $m$ at the position $j$ is denoted by $x_{l,m,j}$, or $x_{l,j}$ when there is only one map in the layer. The same way, we define $\sigma_{l,m,j}$ as the scalar product between a set of input neurons and the weight connection between these neurons and the neuron number $j$ in the map $m$ in the layer $l$

$$x_{l,m,j} = f(\sigma_{l,m,j}) \tag{2}$$

where $f$ can change function to the layer:

- This sigmoid function (hyperbolic tangent) is almost linear between $-1$ and 1, $f(1) = 1$ and $f(-1) = -1$, the constants are set according to the recommendations described in (LeCun et al., 1998b). It is used for $L_1$ and $L_2$, which represent convolutions of the input signal. It allows keeping the variance of the outputs close to 1

$$f(\sigma) = 1.7159\, tanh\left(\frac{2}{3}\sigma\right) \tag{3}$$

- The classical sigmoid function (logistic function) is used for the $L_4$ and $L_5$

$$f(\sigma) = \frac{1}{1 + exp^{-\sigma}} \tag{4}$$

We define $\sigma_{l,m,j}$ for the four layers. $L_3$ does not contain specific connections, thus there is no $\sigma_{3,m,j}$ to calculate. The neuron values of $L_3$ are calculated directly from $L_2$. $L_1$ and $L_2$ are convolutional layers, respectively in the space and time domain. $L_3$, $L_4$ and $L_5$ can be considered as an MLP where $L_3$ is the input layer, $L_4$ is the hidden layer and $L_5$ is the output layer. For $L_1$ and $L_2$, we can notice that each neuron of the map shares the same set of weights. The neurons of these layers are connected to a subset of neurons from the previous layer. Instead of learning one set of weights for each neuron, where the weights depend on the neuron position, the weights are learned independently to their corresponding output neuron.

- For $L_1$:

$$\sigma_{1,m,j} = w_{1,m,0} + \sum_{i=0}^{i<N_{elec}} I_{i,j} w_{1,m,i} \tag{5}$$

where $w_{1,0,j}$ is a threshold. A set of weights $w_{1,m,i}$ with $m$ fixed, $0 \leqslant i < N_{elec}$ corresponds to a spatial filter, i.e. a channel. In this layer, there are $N_{elec} + 1$ weights for each map. For instance, this layer may cancel some artifacts due to difference of the phase in the signal between electrodes. Contrary to the classical definition of linear spacial filters, a threshold is used to be consistent with all the other basic neuron units of the network.

- For $L_2$:

$$\sigma_{2,m,j} = w_{2,m,0} + \sum_{i=0}^{i<N_k} x_{1,m,j-i} w_{2,m,i} \quad (6)$$

where $w_{2,0,j}$ is a threshold. A set of weights $w_{2,m,i}$ with $m$ fixed, $0 \leqslant i < N_k$ correspond to a filter in the time domain. In this layer, there are $N_k + 1$ weights for each map. In the experiments, $N_k$ is fixed to 16. As the sampling rate is 128 Hz, 16 points correspond to 125 ms.

- For $L_4$:

$$\sigma_{4,j} = w_{4,0,j} + \sum_{i=0}^{i<N_s} \sum_{k=0}^{k<N_f} x_{3,i,k} w_{4,i,k} \quad (7)$$

where $w_{4,0,j}$ is a threshold. Each neuron of $L_4$ is connected to each neuron of $L_3$. $L_4$ and $L_3$ are fully connected. In this layer, each neuron has $N_s N_f + 1$ input weights. $L_4$ contains 100 ($N_s N_f$) input connections.

- For $L_5$:

$$\sigma_{5,j} = w_{5,0,j} + \sum_{i=0}^{i<100} x_{4,i} w_{5,i} \quad (8)$$

where $w_{5,0,j}$ is a threshold. Each neuron of $L_5$ is connected to each neuron of $L_4$.

The states of the neuron in $L_3$ are not computed by a classical propagation. They correspond to the results of different signal processing parts. $L_3$ is the result of $L_2$ after the Fourier transform and a selection of specific values. The Fourier transform is applied on $L_2$ in order to pass in the frequency domain as high level features are in this domain. Besides, particular frequencies are chosen in relation to the frequencies of the stimuli.

The value of each neuron of $L_3$ is defined as:

$$x_{3,m,j} = |Y_m(S(j))| \quad (9)$$

where

$$Y_m(u) = \frac{1}{T} \sum_{v=0}^{v<T} y_m(v) exp^{\frac{-2\pi i}{N_t - N_k} uv} \quad (10)$$

with

$$y_m(v) = \begin{cases} x_{2,m,j} & \text{if } v < N_t - N_k \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

and $S$ is a function that associates for each neuron $j$ of a map of $L_3$ a specific value in the frequency domain. This value corresponds to a real frequency, which is determined in relation to the sampling rate and the expected SSVEP responses. These selected frequencies are detailed in Section 4.1. $S$ selects the relevant frequency amplitudes calculated.

$Y_m$ represents the Fourier transform of $y_m$. $Y_m(u)$ is based on $x_{2,m}$ with 1024 points by using zero padding, $T = 1024$. The values in $Y_m$ are not only computed for the $u$ that correspond to $S(j)$, $0 \leqslant j < N_f$. Indeed, the phase must be conserved to reconstruct the signal in the time domain during the backpropagation. We note $\theta_m(u)$ the phase of the transformed signal.

### 3.4. Backpropagation

The learning algorithm for tuning the weights of the network uses the classical backpropagation for the layers $L_5$, $L_4$, $L_3$ and $L_1$. The weights are corrected thanks to a gradient descent by minimizing the least mean square error. For $L_2$, the errors are calculated in a different way as there exists no connection between $L_2$ and $L_3$. For $L_3$, the error must be transferred back in the time domain from the frequency domain by using the Inverse Fourier Transform in order to calculate the errors of $L_2$.

- For $L_5$ the error of each neuron $j$ is defined by:

$$\delta_{5,j} = (o(j) - x_{5,j}) f'(x_{5,j}) \quad (12)$$

where $o(j)$ is the expected value for the neuron $j$.

- For $L_4$, $L_3$ and $L_1$ the error is defined by:

$$\delta_{l,m,j} = f'(x_{l,m,j}) \sum_{i=0}^{N_{out}} w_{l+1,m,i} \delta_{l+1,m,i} \quad (13)$$

where $N_{out}$ is the number of neurons that have $n(l,m,j)$ as input. The weights $w_{l+1,m,i}$ are on the connection between $(l,m,j)$ and the neurons $(l+1,m,i)$.

- For $L_2$, we first define $Z_m$ that represents the error at every frequency:

$$Z_m(v) = \begin{cases} \delta_{3,m,(S^{-1}(v))} exp^{i\theta_m(u)} & \text{if } S^{-1}(v) \text{ is defined} \\ 0 & \text{otherwise} \end{cases}$$

The inverse Fourier transform is applied on $Z_m$:

$$z_m(u) = \sum_{v=0}^{v<T} Z_m(v) exp^{\frac{2\pi i}{N_t - N_k} uv} \quad (14)$$

$z_m$ contains complex values, however the input signal contains only real values. Therefore, we consider only the real part of the error for updating the weights:

$$\delta_{2,m,j} = |z_m(j)| \quad (15)$$

Each weight is updated by $\Delta w_{l,m,i}$:

$$\Delta w_{l,m,i} = \gamma \delta_{l+1,m,j} x_{l,m,i} \quad (16)$$

where $w(l,m,i)$ is the weight on the connection between $n(l,m,i)$ and $n(l+1,m,j)$.

For layers $L_4$ and $L_5$, the learning rate $\gamma$ is defined by:

$$\gamma = \frac{\lambda}{\sqrt{n(l,m,i)_{N_{input}}}} \quad (17)$$

where $n(l,m,i)_{N_{input}}$ is the number of inputs of $n(l,m,i)$ and $\lambda$ is a constant. During the experiments, $\lambda = 0.2$. This parameter is set in relation on prior experiments on different datasets.

For layers $L_1$ and $L_2$, there is a different learning rate that takes into account the weight sharing

$$\gamma = \frac{2\lambda}{n(l,m,0)_{N_{shared}} \sqrt{n(l,m,i)_{N_{input}}}} \quad (18)$$

where $n(l,m,0)_{N_{shared}}$ is the number of neurons that share the same set of weights. For $L_1$ and $L_2$, each neuron on each map shares the same number of weights.

At the initialization of the network, the weights and the thresholds of each neuron are initialized with a standard distribution around $\pm 1/n(l,m,i)_{N_{input}}$.

### 3.5. Evaluation

For the evaluation of the classifier $E$, a pattern $X$ is successfully affected to the class $C_i$, $0 \leqslant i < M$, i.e. $E(X) = C_i$ if

$$i = \arg\max_{0 \leqslant j < M} x_{5,j} \quad (19)$$

The recognition rate $\tau_{rec}$ and the error rate $\tau_{err}$ are defined by:

$$\tau_{rec} = \frac{\sum_{X \in DB} ((E(X) = C_i) \text{ and } (X \in C_i))}{\sum_{X \in DB} X \in C_i} \quad (20)$$

$$\tau_{err} = 1 - \tau_{rec} \quad (21)$$

where $DB$ is a database that contains patterns of size $N_{elec} N_t$.

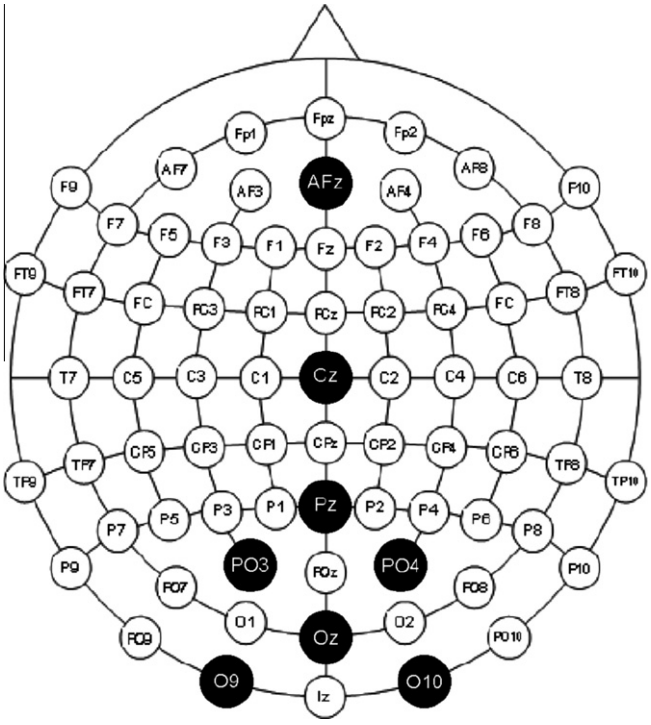


**Fig. 2.** Electrode locations.

## 4. Experiments

EEG data were recorded from the surface of the scalp via eight standard EEG electrodes. They are placed on position $AF_Z$ for ground, $C_Z$ for the reference and $PO_3$, $PO_4$, $P_Z$, $O_9$, $O_{10}$, $O_Z$ for the input electrodes (Chatrian et al., 1985). The location of the electrodes is depicted in Fig. 2. A g.USBamp EEG amplifier from g.tec with a sampling frequency set at 128 Hz was used for the experiments. An analog bandpass filter between 2 and 30 Hz, and a notch filter around 50 Hz (main frequency in Europe) were applied directly inside the amplifier during the EEG acquisition. As the quality of the SSVEP response depends on the stability of the frequencies, the following frequencies are used during the experiments: 6.66, 7.50, 8.57, 10.0 and 12.00 Hz ($N_{freq} = M = 5$).

The protocol experiments was tested on 10 healthy volunteer subjects. The average age of the subjects is 27.2 years, with a standard deviation of 2.44. For each of the five frequencies, six sessions are recorded. A session corresponds to watching during 20 s an isolated stimulus. The task was to focus on the flickering box on the screen. The subjects were instructed to gaze at the flashing targets. Thanks to these sessions and to the different instructions to the subject, the ground truth is reliable and easier to establish. For each subject, we acquire the equivalent of 10 min of EEG ($5 * 6 * 20$ s).

### 4.1. Features selection

We can expect a high peak in the amplitude when a subject focuses on the stimulus flickering at one of the five frequencies ($N_{freq} = 5$). Moreover, their harmonics can also contain information for the classification (Müller-Putz et al., 2005). Table 1 presents for each class, the relevant frequencies for the classification. The number of possible features is 15, but the frequencies 30.00 Hz and 36.00 Hz are considered too high to contain reliable information. Besides, the frequency 20 Hz is present two times, it is the second harmonic of 10 Hz and the third of 6.66 Hz. Therefore, the number of selected frequencies is established at 12, $N_f = 12$.

**Table 1**
Pertinent frequencies for each class.

| $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|
| 6.66 | 7.50 | 8.57 | 10.00 | 12.00 |
| 13.33 | 15.00 | 17.14 | 20.00 | 24.00 |
| 20.00 | 22.50 | 25.71 | 30.00 | 36.00 |

### 4.2. EEG database

The learning, validation and test data set contains respectively 60%, 20% and 20% of the recorded EEG signal. The database was not mixed in relation to the EEG acquisition order. It means that for each frequency, the recorded data in the training set are anterior to the data recorded in the validation, which are anterior to the test set. Data between the test and the training database are totally uncorrelated, as they belong to different sessions. In each database, in each session, each successive pattern overlapped the preceding pattern by 90% when the time segment length is 1 s. Each pattern in the database represents a sliding window on the signal, which is shifted every 100 ms. The learning, validation and test database contain 2700, 900 and 900 patterns, respectively, and for each subject. The selected time segments of 1 s is chosen to highlight the efficiency of the method. It is quite easy to classify SSVEP responses in a long segment length (more than 4 s) with minimum signal pre-processing and only the Fourier transform combined with thresholds. The peaks corresponding to the stimulus frequency can be identified easily; the classification is relatively easy. The detection of an SSVEP response in short time segments like 1 s is a challenge and needs efficient pattern recognition methods. In Trejo et al. (2006), the control lag in a SSVEP-BCI was estimated to 1–5 s. In (Martinez et al., 2007), consider a sliding time window of 4 s. In Wang et al. (2006), an average time of 3.4 s, 4.87 s and 5.58 s are needed for the selection of a command for three subjects.

### 4.3. Complexity

Although the neural network architecture contains several hidden layers, the weights are shared for every neuron within one map in the first two hidden layers. It therefore reduces the number of free parameters in the network. For each layer, the number of parameters (the number of weights and thresholds for all the neurons) is defined as follows:

- In $L_1$, the number of free variables is $N_s(N_{elec} + 1)$, e.g. 42 parameters.
- In $L_2$, the number of free variables is $N_s(N_k + 1)$, e.g. 102 parameters.
- In $L_4$, the number of free parameters is $N_s(N_f + 1) * 100$, e.g. 7272 parameters.
- In $L_5$, the number of free parameters is $M * 101$, e.g. 505 parameters.

where $N_{elec} = 6$ as defined in Section 4, $N_t = 128$, e.g. 1 s as the sampling rate is 128 Hz, $N_f = 12$, $N_s = 6$, $sN_k = 16$ and $M = 5$. The choice of $N_s$ is driven by the further comparisons detailed in the following section. The training time depends on the subject and on the initial learning parameter $\lambda$. The average training time was around 20 min on an Intel Core 2 Duo T7500 CPU. This slow learning is mainly due to the low learning rate $\lambda$. For the test, it is possible to classify about 534 patterns (representing 1 s of signal) per second. Thus, online processing is not an issue. The model was implemented in C++ without any special hardware optimization (multicore or GPU (Meuth and Wunsch, 2007)).

**Table 2**
Recognition rate with a time segment of 1 s.

| Subject | $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ | Min | Max | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0.00 |
| 2 | 99.44 | 100.0 | 96.11 | 95.00 | 92.78 | 92.78 | 100.0 | 96.67 | 3.04 |
| 3 | 98.89 | 98.33 | 98.89 | 86.67 | 82.22 | 82.22 | 98.89 | 93.00 | 7.97 |
| 4 | 96.11 | 99.44 | 92.78 | 95.00 | 99.44 | 92.78 | 99.44 | 96.55 | 2.89 |
| 5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0.00 |
| 6 | 93.89 | 89.44 | 90.56 | 86.11 | 73.33 | 73.33 | 93.89 | 86.67 | 7.96 |
| 7 | 82.22 | 95.56 | 96.67 | 84.44 | 82.78 | 82.22 | 96.67 | 88.33 | 7.16 |
| 8 | 100.0 | 100.0 | 98.89 | 91.67 | 100.0 | 91.67 | 100.0 | 98.11 | 3.63 |
| 9 | 100.0 | 94.44 | 98.89 | 100.0 | 97.22 | 94.44 | 100.0 | 98.11 | 2.35 |
| 10 | 100.0 | 100.0 | 100.0 | 93.89 | 99.44 | 93.89 | 100.0 | 98.67 | 2.68 |
| Min | 82.22 | 89.44 | 90.56 | 84.44 | 73.33 | 73.33 | 93.89 | 86.67 | 0.00 |
| Max | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 7.97 |
| Mean | 97.06 | 97.72 | 97.28 | 93.28 | 92.72 | 90.33 | 98.89 | **95.61** | 3.77 |
| SD | 5.61 | 3.55 | 3.28 | 5.94 | 9.74 | 8.50 | 2.04 | 4.74 | 2.97 |

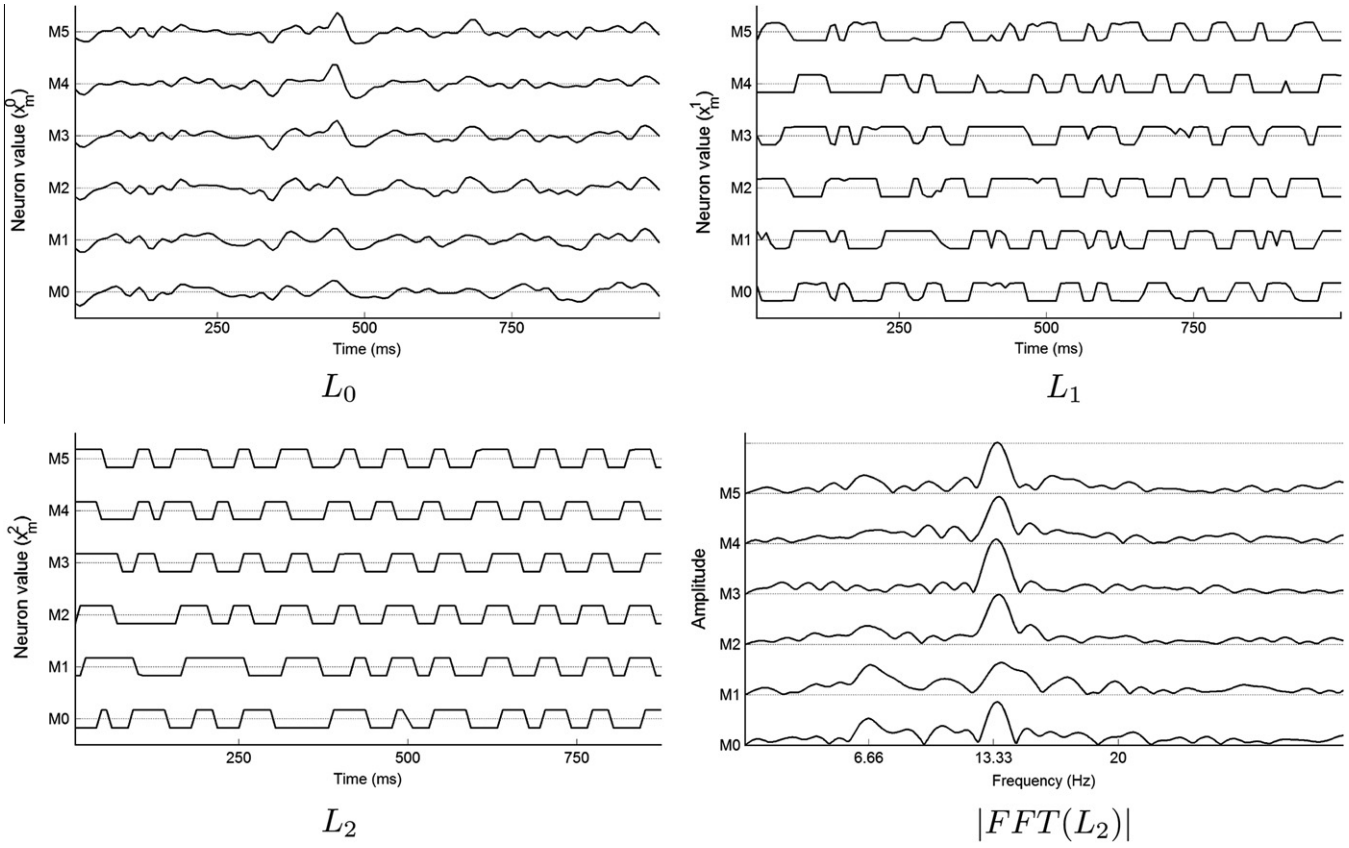


**Fig. 3.** Representation of the neuron values for the input layer, the two first hidden layers, and the Fourier transform of $L_2$ before the frequency selection (subject 1 – $f$=6.66 Hz).

## 5. Results

EEG signals are known for their high variability across subjects. This assumption involves the training of one classifier for each subject. Table 2 presents the recognition rate (in %) for each of the 10 subjects and 5 considered classes $C_i$, $0 \leqslant i < M$. For the frequencies and subjects, the minimum, the maximum, the mean and the standard deviation (SD) are given. As expected, there exists an important variability in the accuracy across subjects and across classes for the same subject. The best results are obtained with subjects 1 and 5 where the accuracy is perfect. For a time segment analysis of 1 s, subject 6 offers the worst recognition rate with a mean of 86.67%. The mean accuracy for the CNN is 95.61%.

Fig. 3 illustrates the behavior of the network for the first two hidden layers on subject 1 once the network has been trained on one pattern, which represents 1 s of the EEG signal. The neuron values of these layers can be interpreted as some signal processing. Each line of $L_0$ represents the EEG input for each electrode after normalization. For $L_1$, $L_2$ and its transformation, each line corresponds to a channel, e.g. the neuron value of a map. The signal to detect in Fig. 3 corresponds to an SSVEP response for a visual stimulus flickering at 6.66 Hz. Although the transformation of $L_2$ is one step towards the classification, it is possible to clearly distinguish a peak at one frequency. $L_2$ has an almost pulse wave shape. This observation is more evident with the signal extracted from subject 1, which is translated in a high recognition rate. During the prop-

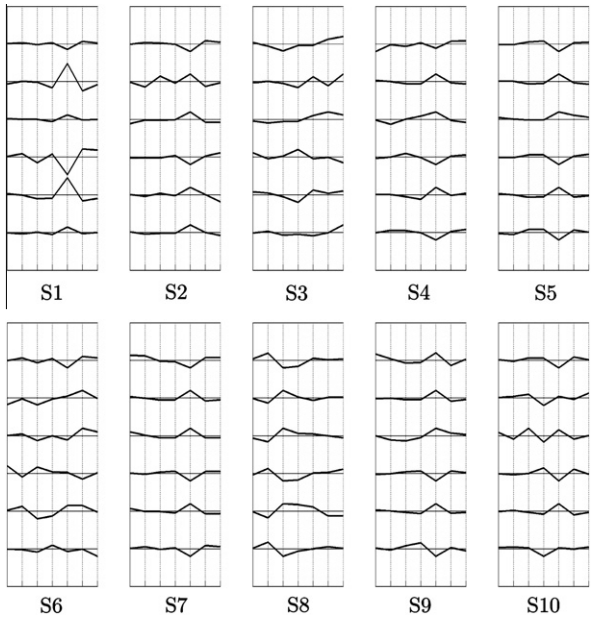


**Fig. 4.** Visual representation of the weights learned in $L_1$ for every subject (the weight semantic denotes spatial filters).

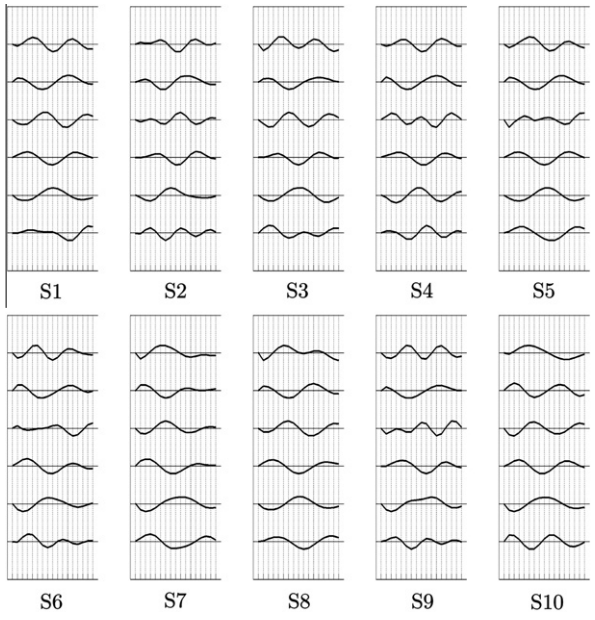


**Fig. 5.** Visual representation of the weights learned in $L_2$ for every subject (the weight semantic denotes temporal filters).

agation, the impact of the time filters seems to be relevant as the shape of the signal clearly change between the $L_1$ and $L_2$.

Fig. 4 presents the weights learned in $L_1$ for each of the 10 subjects (Sn is the $n$th subject, $1 \leqslant n \leqslant 10$), with a window length of 1 s. For each subject, each curve represents the values of the threshold and the weights of the different channels (from bottom to top, $M_0$ to $M_5$). The values of each curve correspond in the order to the threshold, and then the weights corresponding to the electrodes $P_Z$, $PO_3$, $PO_4$, $O_Z$, $O_9$, $O_{10}$ in this order. For instance, we can note a high influence of the electrode $O_Z$ in every channel for subjects 1, 7 and 9. For subject 6, the set of weights seems less homogeneous between channels. The absolute value of the different weights can be interpreted as the discriminant power of the corresponding electrode for the final classification.

Fig. 5 presents the weights learned in $L_2$ for every subject. The first point of each curve corresponds to the threshold. Then, the following 16 points represent the weights in the order over time. As these filters are set in relation to the error in the frequency domain in the upper layers during the learning step, they already preselect the discriminant frequencies. Once the signal is denoised, the Fourier transform is only needed to shift the neural analysis from the time to the frequency domain. In addition to provide to the upper layers directly the discriminant frequencies, the classification is independent of the signal phase by using the amplitudes of the Fourier transform.

### 5.1. Comparisons

To validate the approach, the proposed neural network topology is compared with others. The comparison of the recognition rate for several methods is given in Table 3. Usually, the classical approach is to use the amplitude of the frequencies directly as input of the classifier. Then, the spatial filters can be determined in relation to the frequency powers of some relevant frequencies. To test this strategy, we consider a new convolutional neural network: CNN-F. This network is composed of four layers. The topology of this network is presented in Fig. 6.

- Layer 0 ($L_0$): the input layer. $I_{i,j}$ with $0 \leqslant i < N_{elec}$ and $0 \leqslant j < N_f$. Like for the previous experiment, $N_{elec} = 6$ and $N_f = 12$. The input contains the frequency power for each electrode and for each frequency defined in Section 4.1.
- Layer 1 ($L_1$): the first hidden layer is composed of $N_s$ maps. Each map of $L_1$ has the size $N_f$. In the experiment, $N_s = 6$.
- Layer 2 ($L_2$): the fourth hidden layer is composed of 1 map of 100 neurons. The size of this layer was chosen based on trial runs. This map is fully connected to the different maps of $L_1$.
- Layer 3 ($L_3$): the output layer. This layer has only one map of $M$ neurons, which represents $M$ frequencies to detect. This layer is fully connected to $L_2$.

This neural network was trained and tested in the same condition than the one described in Section 3. The average recognition rate of

**Table 3**
Method comparison for the channel creation.

| Method | Subjects | | | | | | | | | | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | |
| CNN-F | 98.33 | 91.33 | 95.78 | 88.67 | 97.11 | 72.56 | 80.22 | 98.67 | 83.22 | 84.89 | 89.08 | 8.79 |
| CNN + average | 70.44 | 76.11 | 63.67 | 50.56 | 48.11 | 34.89 | 56.56 | 85.33 | 33.00 | 39.67 | 55.83 | 17.84 |
| CNN + native | 95.33 | 87.00 | 89.56 | 70.33 | 79.67 | 53.22 | 68.67 | 94.22 | 67.56 | 70.44 | 77.60 | 13.77 |
| CNN + Laplacian | 99.11 | 93.22 | 89.00 | 94.11 | 100.0 | 77.67 | 79.00 | 90.33 | 95.33 | 95.78 | 91.36 | 7.66 |
| CNN + GD | 100.0 | 96.67 | 93.00 | 96.55 | 100.0 | 86.67 | 88.33 | 98.11 | 98.11 | 98.67 | 95.61 | 4.74 |
| MEC | 99.54 | 93.26 | 81.08 | 90.50 | 99.97 | 75.98 | 75.93 | 98.53 | 98.07 | 90.62 | 90.35 | 9.50 |

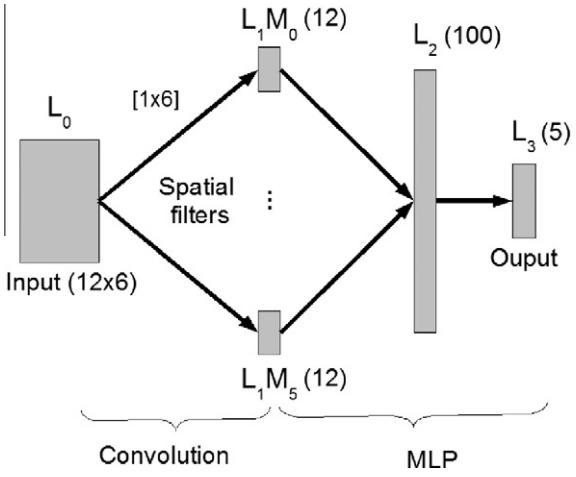


**Fig. 6.** Overview of CNN-F.

CNN-F is 89.08%. It shows that the creation of spatial filters directly in relation to the amplitude of the frequencies is less effective. The accuracy is bellow the proposed strategy. It proves the interest of allowing discriminant steps before using the amplitudes as features.

In the topology of the *CNN*, the number of channels is chosen equal to the number of electrodes in order to facilitate the comparison with other spatial filters ($N_s = N_{elec} = 6$). Three fixed spatial filters are tested and compared. In these methods, the first hidden layer is static: no weights are updated and each neuron has a null threshold. These methods highlight the importance of the creation of spatial filters. Instead of separating the fixed spatial filters from the classifier, they are set directly inside the network to have a rigorous and fair comparison between classical spatial filters and learned filters. The only difference is that the backpropagation will stop at $L_2$. The weights in $L_1$ will not be updated, simulating fixed filters. Furthermore, the chosen activation function in $L_1$ will not change the output range of the fixed spatial filters.

- CNN + average: $L_1$ establishes the average of the electrode values. The weights on the connections between $L_0$ and $L_1$ are equal to $1/N_{elec}$.
- CNN + native: $L_1$ establishes the native combination of the electrode values. For each neuron of the map $i$, $0 \leqslant i < N_{elec}$, the weights on the connections between $L_0$ and $L_1$ are defined by:

$$w(1,m,j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

  with $0 \leqslant j < N_{elec}$.
- CNN + Laplacian: The effect of $L_1$ on input is a Laplacian filter. For each neuron of the map $i$, $0 \leqslant i < N_{elec}$, the weights on the connections between $L_0$ and $L_1$ are defined by:

$$w(1,m,j) = \begin{cases} N_{elec} - 1 & \text{if } j = i \\ -1 & \text{otherwise} \end{cases} \tag{23}$$

  with $0 \leqslant j < N_{elec}$.
- CNN + GD: The weights are established by a gradient descent (GD) as described in Section 3.

The discriminant power of the frequency power for each electrode can be corrected during the training of the neural network with the CNN + native method, contrary to the CNN + average. With the CNN + average solution, each map of $L_2$ contains the same information. The low average recognition rate (55.83%) of this method shows the importance of the spatial filters for specifying for each electrode a specific weight. With the native combination, the recognition rate remains low (77.60%) in spite of the learning procedure. The Laplacian filter is described as a good spatial filter in the literature (Friman et al., 2007) and offers results that can be compared with the learned filters. For the Laplacian method, the average recognition rate is 91.36%. The proposed method where the spatial filters are discovered during the learning of the network proves the interest of the strategy as the average recognition rate is 95.61%. The dichotomy between the methods is respected for most of the subjects. The best improvement is observed for subject 6, where the recognition rate is increased by 9% with the CNN + GD method. For subject 8, the improvement is not as important as the results of the CNN + average solution are already superior to 85%.

Finally, the neural network has been compared with spatial filters obtained with the minimum energy combination (MEC) described in (Friman et al., 2007). The average recognition rate is 90.35% for MEC. The CNN + GD outperforms the MEC method with a difference of 5.26%. The recognition rate for each class is given in Table 3.

A statistical analysis has been performed to compare the performance of the best classifiers across subjects. A pairwise one tail t-test comparison indicates that CNN + GD is superior to CNN-F ($p = 0.0048$, $t_9 = 3.269$, SD = 6.320). With the same type of comparison, CNN + GD is superior to CNN + Laplacian ($p = 0.0014$, $t_9 = 4.075$, SD = 3.3027). Finally, CNN + GD is also superior to MEC ($p = 0.0056$, $t_9 = 3.1798$, SD = 5.2340). $t_9$ and SD are the value of the test statistic with the degrees of freedom of the test (9), and the sample standard deviation, respectively. This analysis proves the high performance of CNN + GD across subjects.

## 6. Discussion

Experiments conducted on different trials have proven the efficiency of this new CNN architecture. The method outperforms the traditional way that uses the amplitude of the frequencies directly as input. It also outperforms fixed spatial filters. The channel creation is a crucial aspect for tailoring a classifier for EEG. A large difference exists between the raw average of the electrodes and advanced techniques for creating spatial filters. The recognition rate obtained with a time segment of 1 s proves the relevance of the method for the detection of SSVEP responses. Such method could be used in real BCI applications. In spite of the variability between the subjects, the classifier is able to provide a high accuracy for each subject.

The high accuracy of the SSVEP classification shall be weighted for its interpretation during an online BCI experiment. As explained in the protocol section, it is difficult to evaluate the raw classification procedure as the exact ground truth is difficult to create. This paper focused on the classification of well identified brain responses. Although the classification of particular brain signals and the command detection in a BCI are directly related, they represent different measurements. In addition, asynchronous BCI assumes an $N_{freq} + 1$ classes, $N_{freq}$ visual frequencies and one class that represents the moment where the subject is not looking at any frequencies.

The proposed method determines appropriate spatial filters that optimize the discriminant power of multi-channel brain signals based on SSVEP responses for the classification. They dictate the quality of the classification as suggested by the comparison with other methods to set the channels. The choice of the ideal number of channels, i.e. spatial filters in the neural network topology remains a pre-defined choice. A higher number of channels, i.e. a higher number of maps in the network, could translate more fluctuations of the brain activity. However, optimal spatial filters may vary over time for some subjects. In such case, the accuracy will vary and the network should be trained regularly to keep optimal channels. The system could also be invariant to non-task-related fluctuations in the EEG signal during feedback. These fluctuations can be caused

by changes in the subject's brain processes. It can be due to the fatigue or the involvement of other tasks. Muscular artifacts coming from swallowing or yawning may also involve other fluctuations in the signal (Blankertz et al., 2008a). Two ways are possible for keeping a stable reliability over time. First, the system can be adapted regularly in an incremental way. Second, the system shall become invariant to the signal fluctuations. The choice of the approach can depend on the nature of the fluctuations.

Two main research directions could improve the quality of the current classifier. First, the current system integrates only the Fourier transform as an external transformation. Further works will deal with other methods like wavelets, to add new knowledge and directions about the neural processing. Second, the accuracy represents the recognition rate of the classifier without rejection. Future studies will investigate the integration of rejection criteria that will improve the reliability of the classification and its integration in BCI applications.

## 7. Conclusion

A new convolutional neural network has been presented. It includes one particular novelty: the semantic of the neuron values is changed from the time domain to the frequency domain thanks to the Fourier transform. The transformation is inserted between two hidden layers. Input features are first processed in the spatial domain, then in the time domain and finally in the frequency domain. While the neural network can be considered as a black box, its topology allows the analysis of the best spatial and time filters needed for the classification. Compared with other neural network topologies, the proposed classifier gives the best accuracy. The efficiency of the technique has been proven on the offline classification of a popular brain response, which can be used in Brain–Computer Interfaces.

## Acknowledgments

This research was supported by a Marie Curie European Transfer of Knowledge grant BrainRobot, MTKD-CT-2004-014211, within the 6th European Community Framework Program. The data were collected at the Institute of Automation, University of Bremen, Bremen, Germany.

## References

Anderson, C.W., Devulapalli, S.V., Stolz, E.A., 1995. Determining mental state from EEG signals using parallel implementations of neural networks. In: IEEE Workshop on Neural Networks for Signal in Processing, Cambridge, MA, USA, pp. 475–483.

Barreto, A.B., Taberner, A.M., Vicente, L.M., 1996. Neural network classification of spatio-temporal EEG readiness potentials. In: Proc. Fifteenth Southern Biomedical Engineering Conf., pp. 73–76.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., 2007. Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems, vol. 19. MIT Press.

Blankertz, B., Curio, G., Müller, K.R., 2002. Classifying single trial EEG: Towards brain computer interfacing. In: Diettrich, T.G., Becker, S., Ghahramani, Z. (Eds.), Advances in Neural Inf. Proc. Systems (NIPS 01), vol. 14, pp. 157–164.

Blankertz, B., Dornhege, G., Lemm, S., Krauledat, M., Curio, G., Müller, K.R., 2006. The Berlin brain–computer interface: EEG-based communication without subject training. IEEE Trans. Neural Systems Rehab. Eng. 14, 147–152.

Blankertz, B., Kawanabe, M., Tomioka, R., Hohlefeld, F., Nikulin, V., Müller, K.R., 2008a. Invariant common spatial patterns: Alleviating nonstationarities in brain–computer interfacing. Adv. Neural Inform. Process. Systems, 20.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.R., 2008b. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Proc. Mag. 25, 41–56.

Brunner, C., Naeem, M., Leeb, R., Graimann, B., Pfurtscheller, G., 2007. Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. Pattern Recognition Lett. 28, 957–964.

Burkitt, G., Silberstein, R., Cadush, P., Wood, A., 2000. Steady-state visual evoked potentials and travelling waves. Clin. Neurophysiol. 111, 246–258.

Cecotti, H., Gräser, A., 2008. Convolutional neural network with embedded Fourier transform for EEG classification. In: Proc. 19th Internat. Conf. on Pattern Recognition.

Chatrian, G.E., Lettich, E., Nelson, P.L., 1985. Ten percent electrode system for topographic studies of spontaneous and evoked eeg activity. Am. J. EEG Technol. 25, 83–92.

Felzer, T., Freisieben, B., 2003. Analyzing EEG signals using the probability estimating guarded neural classifier. IEEE Trans. Neural Systems Rehab. Eng., 11.

Friman, O., Volosyak, I., Gräser, A., 2007. Multiple channel detection of steady-state visual evoked potentials for brain–computer interfaces. IEEE Trans. Biomed. Eng. 54, 742–750.

Haselsteiner, E., Pfurtscheller, G., 2000. Using time dependent neural networks for EEG classification. IEEE Trans. Rehab. Eng. 8, 457–463.

Hinton, G., Osindero, S., Teh, Y., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554.

Hinton, G., Salakhutdinov, R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998a. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.

LeCun, Y., Bottou, L., Orr, G., Müller, K.R., 1998b. Efficient backprop. In: Orr, G., Muller, K., (Eds.), Neural Networks: Tricks of the Trade.

LeCun, Y., Huang, F.J., Bottou, L., 2004. Learning methods for generic object recognition with invariance to pose and lighting. In: Proc. CVPR'04. IEEE Press

Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B., 2007. A review of classification algorithms for EEG-based brain–computer interfaces. J. Neural Eng. 4, R1–R13.

Lüth, T., Ojdanic, D., Friman, O., Prenzel, O., Gräser, A., 2007. Low level control in a semi-autonomous rehabilitation robotic system via a brain–computer interface. In: ICORR 2007. IEEE 10th Internat. Conf. on Rehabilitation Robotics, pp. 721–728.

Martinez, P., Bakardjian, H., Cichocki, A., 2007. Fully online multicommand brain–computer interface with visual neurofeedback using SSVEP paradigm. Comput. Intell. Neurosci..

Masic, N., Pfurtscheller, G., Flotzinger, D., 1995. Neural network-based predictions of hand movements using simulated and real EEG data. Neurocomputing 7, 259–274.

Meuth, R.J., Wunsch, D.C., 2007. Approximate dynamic programming and neural networks on game hardware. In: Proc. Internat. Joint Conf. on Neural Networks.

Müller, K.R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B., 2008. Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. J. Neurosci. Methods 167, 82–90.

Müller-Putz, G.R., Pfurtscheller, G., 2008. Control of an electrical prosthesis with an SSVEP-based BCI. IEEE Trans. Biomed. Eng. 55, 361–362.

Müller-Putz, G.R., Scherer, R., Brauneis, C., Pfurtscheller, G., 2005. Steady-state visual evoked potential (SSVEP)-based communication: Impact of harmonic frequency components. J. Neural Eng. 2, 123–130.

Obermaier, B., Guger, C., Neuper, C., Pfurtscheller, G., 2001. Hidden markov models for online classification of single trial EEG data. Pattern Recognition Lett. 22, 1299–1309.

Pfurtscheller, G., Guger, C., Ramoser, H., 1999. EEG-based brain–computer interface using subject-specific spatial filters. In: Internat. Work-Conf. on Artificial and Natural Neural Networks, vol. 2, pp. 248–254.

Rakotomamonjy, A., Guigue, V., 2008. BCI competition iii: Dataset ii – Ensemble of SVMs for BCI p300 speller. IEEE Trans. Biomed. Eng. 55, 1147–1154.

Sejnowski, T.J., Dornhege, G., Millán, J.d.R., Hinterberger, T., McFarland, D.J., Müller, K.R., 2007. Toward Brain–Computer Interfacing (Neural Information Processing). The MIT Press.

Simard, P.Y., Steinkraus, D., Platt, J.C., 2003. Best practices for convolutional neural networks applied to visual document analysis. In: 7th Internat. Conf. on Document Analysis and Recognition, pp. 958–962.

Tomioka, R., Hill, N.J., Blankertz, B., Aihara, K., 2006. Adapting spatial filter methods for nonstationary BCIs. In: Workshop on Information-Based Induction Sciences (IBIS), p. 6.

Trejo, L.J., Rosipal, R., Matthews, B., 2006. Brain–computer interfaces for 1-D and 2-D cursor control: Designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. IEEE Trans. Neural Systems Rehab. Eng., 14.

Wang, Y., Wang, R., Gao, X., Hong, B., Gao, S., 2006. A practical VEP-based brain–computer interface. IEEE Trans. Neural Systems Rehab. Eng., 14.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., 2002. Brain–computer interfaces for communication and control. Clin. Neurophysiol. 113, 767–791.

Zhong, S., Gosh, J., 2002. HMMs and coupled HMMs for multi-channel EEG classification. In: Proc. IEEE Int. Joint Conf. on Neural Networks, vol. 2, pp. 1154–1159.